

Chance Encounters: A First Course in Data Analysis and Inference.

Christopher J. WILD and George A. F SEBER. New York: Wiley, 2000, xviii + 611 pp., \$93.95(H), ISBN: 0-4713-2936-3.

This book is a distinctive addition to the range of elementary textbooks on the market. It belongs to the family of which Moore and McCabe (1999) is a pioneering representative, in that it is focused on data and on statistical reasoning, expecting only a modest level of algebraic or arithmetic skill from a reader. However, in several ways it offers an attractive alternative approach.

The authors distinguish their book as data oriented, using modern exploratory ideas and real data; computer oriented, using several popular software packages to produce analyses; intuitive and discursive without being long-winded, treating probability in a way that avoids the use of complex rules by embedding the ideas in a study of two-way tables; and using statistics to find answers to questions, so that statistical thinking rather than mechanical manipulation is the focus of presentation and of exercises. How well have the authors achieved these aims?

The examples in the text, and the exercises, offer real data for analysis (frequently with sources referenced although some are unpublished), demonstrating the relevance of the material in everyday life. The examples produced come from all over the world, although the authors' residence in New Zealand means that there are more examples from studies carried out in that country than one would usually find. The contexts of the examples are of universal appeal and could mostly have been located anywhere. Measurements in both metric and imperial units appear; this offers opportunities for exercises in scale change as well as reminding those of us from metric societies how complicated imperial units were! The contexts producing the datasets tend to be issues of general interest for which undergraduate students might be expected to have some background knowledge (human behavior, simple economic issues, sport, popular science). There is not a pig diet in sight, although the standard blocked agricultural experiments appear briefly.

Some recent textbooks emphasized the use of published scientific research data as a motivating force to instill in students the idea that the material they are studying is used by serious investigators. However, a frequent consequence of this is that the context of the exercises is at a scientific level that students, unless already advanced in that particular field of study, cannot master sufficiently to exercise "statistical thinking" about the data. Any statistical consultant realizes that acquiring mastery of quite a bit of detail about the field precedes carrying out a successful analysis: so the strategy of presenting students with general interest contexts, rather than specialized scientific examples, seems likely to enable them to engage with the data in a useful way. It is the reviewer's experience that faculty rather than students of service courses are most keen to have specialized examples.

Exercises or examples are intended to be carried out using software; the book is not prescriptive, but gives instructions for using Minitab, Excel, S-Plus, and R, and occasionally also SAS and SPSS. The idea is that a student who understands what is being done will have no difficulty with the menu choices in most menu-driven languages, and explanations are given of major differences between packages for important procedures. This makes the book potentially very flexible to use. Inclusion of references to the free package R makes advanced material accessible to groups with limited financial resources, while instructions and caveats for using Excel

makes it accessible to the vast numbers who have this spreadsheet program.

The authors' intuitive approach to data analysis is similar to that of Moore and McCabe (1999) in its focus on conceptual understanding, on the use of graphical and summary displays, and on the importance of the data gathering process to the eventual possible analyses. The order of presentation is reasonably standard: collection of data (with considerable emphasis on surveys), data exploration for univariate and bivariate data, probability, discrete and continuous random variables, sampling distributions, confidence intervals for means and proportions, significance test concepts. These ideas are then applied in chapters on inference for continuous variables, tables of counts, relationships between quantitative variables. Chapters on Control Charts and on Time Series are available from the authors' Web site.

This book is most distinguished from others by its treatment of probability through two-way and higher-way tables. The authors present on their Web site alternative traditionally based chapters on probability and random variables which could be adopted by those who prefer, but the table-based approach allows for all the relevant probability material to be presented, including expected values, conditional probability, change of measurement scale, and standard deviation. Changing the order of conditioning is explained and a number of interesting problems are presented, without any reference to the Bayes's Theorem algebra. It is very clear, when the given data are entered in the appropriate two-way table, what information is required to solve such problems and how the solution is to be obtained. I look forward to using this approach since my experience with the traditional Bayes's Theorem approach is that students can learn the formula but don't know which bits of information belong where. The extension from the basic probability ideas to discrete and continuous random variables is lucid and intuitively appealing.

Of the other chapters, I found those on exploratory data analysis to be least compelling. Terms are introduced without being explicitly defined, the language level is quite sophisticated: this is a book that students with marginal English skills might find heavy going. Some sections in this early part of the book are quite brief, raising issues and not developing them: the presence of an instructor seems presupposed here.

Later, however, once probability and inference are introduced, the explanations of tricky aspects of the concepts become much more complete. The authors use both verbal and pictorial displays in many cases. They show, for example, finding probabilities $P(a < X < b)$ from the Normal distribution using the usual figures with shaded areas under bell curves, but they also indicate which areas are found from a computer program or table) and how to set up the algebra to find the required result. A student with little mathematical sophistication should be able to use the later part of the book reasonably successfully for self-study, such is the level of detail of the explanations.

The book has several structural elements (including sequence of presentation and types of examples used) designed to enhance student learning. Concepts are introduced in a heuristic way in the initial chapters, and then revisited later with greater development. For example, the two-way table recurs several times: first as a simple display, then as a basis for developing ideas of probability, then later again in terms of models for independence and association. The authors specifically recommend rereading earlier material in key places, as well as revisiting earlier examples: it is this explicit meta-comment on the presentation which

should help students realize that the material is being built up sequentially. Without meta-comments, even quite able students are surprisingly capable of moving from one part to another of a carefully crafted question or instructional sequence without realizing that they should be using the result of part (a) in answering part (b)!

Each section is followed by a quiz, with questions whose answers can be found by reading the material. A small number of exercises using the concepts and techniques of the section are provided for practice; this is not a book that provides a large number of exercises for drill. Each chapter, however, concludes with a larger number of review exercises, which require integration of the material of the chapter and, as the book proceeds, of earlier chapters as well. These are longer, and provide opportunities for carefully exploring the material. I would be very happy to use them for group exercises or as the basis for discussions since they require reflection and thought rather than just straight application of techniques. In the course of reviewing the book I became quite absorbed in reading review exercises in chapter after chapter; they were a refreshing change from textbook exercises where each new context or data set is just a wrapping for numbers about which the same questions are asked each time ("give a 95% confidence interval for ..."). The addition of "what follow-up investigations would you like to see done?" opens up many examples for discussion of assumptions and scientific process.

Another attractive feature of this book is the use of references to other published statistical material for more in-depth information about issues which are raised but not pursued to completion, for example the various aspects of nonsampling error in surveys. This not only gives the student (or the instructor!) access to more information but makes it very clear that the material being presented sits in a wider context. This is in addition to references for sources of data for examples.

Some trivial complaints: figures on the page ahead of the text referring to them seem to appear more often than is quite comfortable, interrupting the flow of reading to turn the page looking for the subject of the discussion. Use of boxes and highlighting for display of items appears unpatterned and inconsistent; not that items highlighted are not usually worth the emphasis, but it is hard to see why other items do not also merit it. However, on the whole the book is attractively designed and presented, and weighing in at three pounds is more compact than many competitors.

Any instructor responsible for teaching statistical thinking should look at this book; its approach and ideas and exercises are refreshing and stimulating, and it says a great many things that I wish all of my students would take to heart!

Margaret MACKISACK
University of Technology, Sydney

REFERENCES

Moore, D. S.. and McCabe, G.E (1999), Introduction to the Practice of Statistics. New York: W.H. Freeman & Co.

The American Statistician, August 2001, Vol. 55, No. 3, P257